Research Paper ■

Comparison of Methodologies for Calculating Quality Measures Based on Administrative Data versus Clinical Data from an Electronic Health Record System: Implications for Performance Measures

PAUL C. TANG, MD, MS, MARY RALSTON, PhD, MICHELLE FERNANDEZ ARRIGOTTI, MPH, LUBNA QURESHI, MS, JUSTIN GRAHAM, MD, MS

A b s t r a c t   New reimbursement policies and pay-for-performance programs to reward providers for producing better outcomes are proliferating. Although electronic health record (EHR) systems could provide essential clinical data upon which to base quality measures, most metrics in use were derived from administrative claims data. We compared commonly used quality measures calculated from administrative data to those derived from clinical data in an EHR based on a random sample of 125 charts of Medicare patients with diabetes. Using standard definitions based on administrative data (which require two visits with an encounter diagnosis of diabetes during the measurement period), only 75% of diabetics determined by manually reviewing the EHR (the gold standard) were identified. In contrast, 97% of diabetics were identified using coded information in the EHR.

The discrepancies in identified patients resulted in statistically significant differences in the quality measures for frequency of HbA1c testing, control of blood pressure, frequency of testing for urine protein, and frequency of eye exams for diabetic patients. New development of standardized quality measures should shift from claims-based measures to clinically based measures that can be derived from coded information in an EHR. Using data from EHRs will also leverage their clinical content without adding burden to the care process.

■ **J Am Med Inform Assoc.** 2007;14:10–15. DOI 10.1197/jamia.M2198.

Background

Since the Institute of Medicine's (IOM's) 2001 call to "cross the quality chasm,"[1] many major health systems have put in place programs to improve the quality of health care in America.[2] To support these quality improvement efforts, there has been a major push to promote the adoption and use of electronic health record (EHR) systems by clinicians,[3]

and personal health record systems by patients.[4] Larger physician practices have made significant strides in deploying EHR systems, but penetration of EHR use among smaller practices lags behind.[5]

Major payers are using payment incentives to motivate providers to demonstrate that they have achieved improved quality.[6] Increasingly, providers are asked to submit process and outcomes data—using different data definitions and different reporting formats—to be used in pay-for-performance programs, quality improvement initiatives, and other public-reporting endeavors. At the heart of any improvement activity must be accurate, reliable, standardized, and cost-effective means for measuring current performance and for setting desired performance goals. While the number of quality measures in use has increased substantially over recent years, debate is surfacing as to whether these diverse measures of quality have actually led to improvements.[6] Some assert that only modest improvements have been achieved since the IOM's *Quality Chasm* report was released.[7]

To be useful, a quality measure should be precisely defined, tied causally to an outcome, and affected by processes that the providers and/or the patients control. The accuracy and validity of the data used to calculate a measure's value are primarily determined by the match between the purpose for which the data was entered (whether on paper or in an electronic system) and the meaning ascribed to that data element when generating a report. The meaning of the data

as entered should match the way it is being used in the quality measure. Relying on administrative data from billing systems to deduce clinical context violates this principle and often produces misleading results that may lead to misinformed policymaking. Studies comparing the clinical accuracy of claims data against data from clinical databases populated by clinicians have documented significant disparities between data collected for different purposes.[8–12] One study involving cardiac diseases compared diagnosis codes, entered by trained medical records professionals into an administrative database for billing purposes, with diagnoses entered by cardiologists in a clinical database in the course of providing care.[8] The agreement rates between the two databases varied from as low as 0.09 to a maximum of 0.83. The authors concluded that claims data were not as useful as clinical data for identifying "clinically relevant patient groups."

It is critically important that a measure accurately identify the target population (i.e., the denominator in quality and performance measures). Yet, despite the known limitations of claims data, the most commonly used algorithms for the identification of target populations rely on claims-based encounter diagnoses. In order to understand both the benefits and limitations of using EHR data, it is important to understand whether there are significant differences between data captured by clinicians in an EHR system and data entered on claims for payment. To date, there has been limited information published on this subject. In one study, use of an EHR-based strategy to identify encounters dealing with pharyngitis was significantly more sensitive than use of administrative data (96% vs. 62%, p < 0.0001).[13] The total super set of charts reviewed, however, consisted only of those records which were either identified by the EHR-based strategy or the administrative data-based strategy, which consequently precludes calculating the sensitivity for either method against a gold standard. Furthermore, analysis of the EHR had to be conducted by searching text strings in notes instead of using coded data because billing codes were recorded on paper forms that were later entered into a separate billing database. Other investigators have also noted the poor sensitivity of claims data (0.60–0.84),[14] but did not compare it to coded information from an EHR. Another group tried to improve the positive predictive value (PPV) of definitions for specific diseases by combining medication data with claims data.[15] Although they were able to iteratively refine their definition to achieve a high PPV for diabetes, they reviewed only a limited number of charts, and did not assess the sensitivity of their method to identify the target population.

In our study, we explored the implications of using traditional claims-based methods to identify patients with diabetes, compared with patient populations identified through expert chart review and with those identified using data from coded fields in an EHR. In addition, we examined the effect of these differences on selected performance measures.

The study was conducted by the Palo Alto Medical Foundation (the clinic), a large multi-specialty group practice in the San Francisco Bay area, and Lumetra, California's quality improvement organization, as part of the Doctor's Office Quality (DOQ) project, under contract by the Centers for Medicare & Medicaid Services (CMS).

Methods

The study population consisted of Medicare fee-for-service (FFS) patients with at least two encounters for any reason at the clinic during a 15-month data period from April 1, 2004, through June 30, 2005.

From the population of Medicare patients defined above, a random sample of electronic patient charts was selected for review to identify patients who met specific criteria for having diabetes mellitus, and to determine quality measures for these patient populations. The study received IRB exemption as a medical records study.

Gold Standard Medical Record Review

Four trained non-physician reviewers analyzed charts randomly drawn from the general Medicare population described above using predefined inclusion criteria. The reviewers sequentially read charts drawn from the randomized pool until 125 diabetic patients were identified. If one of the non-physician reviewers had a question when determining the diagnosis or abstracting the measures, a physician reviewer was consulted to make the final determination.

To identify patients with diabetes, the expert reviewer first checked the problem list for an indication that the sampled patient had diabetes; finding none, the medication list was reviewed for evidence of a prescription for antidiabetic medications; finding none, the lab results were inspected for two or more results suggesting uncontrolled diabetes; finding none, the free-text notes were reviewed for evidence that the sampled patient had diabetes (such as a notation of a diagnosis of diabetes, a prescription for antidiabetic medication, or treatment for a diabetes-related condition such as diabetic foot ulcer); finding none, the reviewer concluded that the sampled patient did not have diabetes. The reviewer conducted the review in the order specified above, stopped at the first place the diagnosis was found, and abstracted the data required to compute the DOQ diabetes measures. The reviewer recorded the data that was used for the positive diagnosis of diabetes on a data collection sheet designed for this study. The diagnosis determined by expert review formed the gold standard in this study.

Queries of the Electronic Health Record System

Electronic queries of coded data in the EHR database for patients meeting specific criteria for diabetes were performed. The criteria used were the same as those used in the gold standard review, with the exception of review of the free-text notes. They included queries for the presence of diabetes on the problem list, use of anti-diabetic medications, and laboratory test results consistent with uncontrolled diabetes. Since computers cannot reliably interpret non-coded free-text data, those criteria were not used in the electronic queries.

Quality Measures for Diabetes Mellitus

The authors adopted the quality measures for diabetes mellitus (DM1-DM8, see Table 1) that were developed for the DOQ project. The diabetes measures were originally developed by the National Diabetes Quality Improvement Alliance (Alliance) and endorsed by the National Quality

Table 1 ■ DOQ Diabetes Mellitus Measures

	Measures	Description
DM-1	HbA1c Management	Percentage of patients with one or more A1c test(s)
DM-2	HbA1c Management Control (measure of poor control)	Percentage of patients with most recent A1c level > 9.0% (poor control)
DM-3	Blood Pressure Control	Percentage of patients with most recent BP < 140/90 mm Hg
DM-4	Lipid Measurement	Percentage of patients with at least one low-density lipoprotein (LDL) cholesterol test
DM-5	LDL Cholesterol Level	Percentage of patients with most recent LDL cholesterol < 130 mg/dl
DM-6	Urine Protein Testing	Percentage of patients with at least one test for microalbumin during the measurement year, or who had evidence of medical attention for existing nephropathy (diagnosis of nephropathy or documentation of microalbuminuria or albuminuria)
DM-7	Eye Exam	Percentage of patients who received a dilated eye exam or evaluation of retinal photographs by an optometrist or ophthalmologist during the reporting year, or during the prior year if patient is at low risk for retinopathy. A patient is considered low risk if all three of the following criteria are met: (1) the patient is not taking insulin; (2) has an A1c < 8%; and (3) has no evidence of retinopathy in the prior year
DM-8	Foot Exam	Percentage of eligible patients receiving at least one complete foot exam (visual inspection, sensory exam with monofilament, and pulse exam)

Forum (NQF). Subsequent to their adoption by the DOQ project, these measures have undergone further refinement and later versions have also been endorsed by the NQF. As a part of the DOQ project, the Iowa Foundation for Medical Care developed the clinical specifications for the performance measures, identified the necessary data elements and codes, and produced a data collection tool for retrospective capture of the measures by expert review. Encounter diagnoses included all ICD-9 codes that were submitted for a claim. Diagnoses associated with procedures ordered during a visit are automatically included as encounter diagnoses.

Sets of measures (claims-based and EHR data-based) were then computed separately for target populations identified through expert review. In the calculation of diabetes quality measures we included all patients over 18 years of age.

Standard techniques were used to compute prevalence of diabetes using each of these methods. Cross tabulations were performed on selected combinations of methods. Fisher's exact test was used to test for significance when comparing the measures in patients with two or more visits for diabetes with those for patients with fewer than two visits for diabetes.

Results
Of the 5,828 FFS Medicare patients identified through claims, 4,635 had at least two billed visits at the clinic for any reason during the measurement period. Of these 4,635

patients, 17.2% (795) had at least one visit for diabetes, and 14.3% (663) had at least two visits for diabetes. Three percent (132) of patients had only one billed visit for diabetes. The average age for each of the various subsets of patients was similar, around 74 years of age.

The reviewers examined 818 randomly assigned charts to identify 125 (15.3% of the study population) diabetics (see Table 2). There were three charts of patients who met the electronic criteria for diabetes mellitus, but whose diagnosis of diabetes mellitus was not confirmed through the initial gold standard manual review. A second manual review was conducted for quality check two months later; in two of these records, the diagnosis of diabetes mellitus was found in the problem list, but in the third record diabetes mellitus was still not found. This discrepancy may have occurred because problems can be updated between the time of the expert review and the time the electronic query was performed.

Electronic coded data in the problem list, medication list, and lab results in the reviewed charts were compared to the findings of the expert reviewers in Table 3. Diabetes was most often identified through a specific diagnosis in the problem list on the EHR. Additional patients with diabetes were found through coded data in the medication list or laboratory results, and through other information contained in free-text fields, such as progress notes, in the EHR. Using coded information in the EHR, the sensitivity for detecting diabetes was 97.6%, and the specificity was 99.6%. For many

Table 2 ■ Identification of Patients with Diabetes through Electronic Data Capture Compared to the Gold Standard Diagnosis of Expert Review

Electronic Capture of EHR Data Elements		Gold Standard		
		Yes	No	total
Any electronic capture*	Yes	122	3	125
	No	3	690	693
All charts reviewed		125	693	818

*Diabetes on problem list, antidiabetic medication, or two laboratory tests consistent with diabetes.

Table 3 ■ Identification of Patients with Diabetes through Manual Abstraction by Expert Reviewers (Gold Standard)

Diagnosis of Diabetes Determined by	Number of Patients (%)
1. Diagnosis found on problem list	116 (93%)
2. Antidiabetic medication on medication list	4 (3.2%)
3. Two lab tests consistent with diabetes	1 (0.8%)
4. Evidence found in free text fields	4 (3.2%)
Total with diabetes	125

Table 4 ■ Confirmation of Diabetes Compared to Number of Visits with Diabetes Mellitus as an Encounter Code

Method of Confirmation	Result	Number of Visits with DM as Encounter Code			
		2 or More	1	0	Total
Expert review	Yes	94	17	14	125
	No	0	0	693	693
Electronic capture from problem list	Yes	93	16	8	117
	No	1	1	699	701
Electronic capture from problem list, medications,	Yes	94	16	12	122
or laboratory results	No	0	1	695	696
All charts reviewed		94	17	707	818

of the patients, diabetes could have been identified through more than one type of coded data, but only the first data source is included in Table 3. Using coded data from the EHR, 97% of diabetics determined by manual chart review were correctly identified.

Standard DOQ quality measures require that the patient have at least two billed visits for diabetes during the measurement period to be included in the denominator. As shown in Table 4, only 75% (94 of 125) of patients identified as having diabetes by expert review met this requirement. The other 25% (31 of 125) of patients identified as having diabetes by expert review had zero or one visit with diabetes as an encounter diagnosis. All of the patients with one visit for diabetes, and 14 additional patients with no visits for diabetes, were determined to have diabetes by expert review. Similar results were found when electronic data capture was used to identify patients having diabetes.

In Table 5, the measures computed for patients with two visits for diabetes are compared to those for patients with less than two visits, using data determined through expert review. The difference between these two groups is statistically significant for DM1—HbA1c Testing (97% vs. 68%, p < 0.001), DM3—Blood Pressure Control (61% vs. 45%, p = 0.05), DM6—Urine Protein Testing (85% vs. 55%, p < 0.001), and DM7—Eye Exam (62% vs. 41%, p = 0.03). The other measures showed differences that did not achieve statistical significance due to the small sample size.

Discussion

As the country intensifies its efforts to standardize performance measures, and payers offer incentives to providers to adopt EHR systems, it is timely and important to assess the validity of claims data to identify populations of interest and consequently to serve as the basis for measures used in quality improvement and public reporting. In the past, claims data were often used because no other data sources were readily available for large-scale analysis. Consequently, many of the existing methods for identifying a target patient population were created to accommodate limitations of claims data. However, in the absence of true clinical data, it can be very hard to identify which patients have a condition of interest from claims data alone.

Commonly used administrative measures include a requirement that a patient have two visits with an encounter diagnosis of interest. The two-visit requirement probably was intended to assure a certain level of specificity in determining whether a patient has the target condition.[16] Unfortunately, the requirement may disqualify a significant number of patients who have the disease. In our study, we found that requiring two claims with diabetes as an encounter code excluded 25% of patients found to have diabetes by expert review, despite each patient having had multiple visits to the practice during the measurement period. Diabetes may not have been entered as an encounter diagnosis either because it was not dealt with during a specific visit or the person entering the billing diagnosis may not have entered all the diagnoses associated with that encounter.

In contrast, patients with a specific diagnosis can often be easily identified using coded data in an EHR, such as an entry on the problem list. The ability to reliably identify patients based on diagnoses listed in the problem list depends on the use of standardized codes for diagnoses.

Table 5 ■ DOQ Diabetes Measures Calculated from Expert Review Data for all Patients Identified as Having Diabetes: Comparison of Patients with Two Visits for Diabetes Vs. Patients with Zero or One Visit

Measure (Probability, Fisher's Exact Test)	Zero or One Visit for Diabetes		Two visits for Diabetes	
	N (%)*	D*	N (%)*	D*
DM1: HbA1c Management (p < 0.001)	21 (68%)	31	91 (97%)	94
DM2: HbAlc Management Control (measure of poor control) (p = 0.27)	0 (0.0%)	21	6 (6.6%)	91
DM3: Blood Pressure Control (p = 0.05)	14 (45%)	31	57 (61%)	94
DM4: Lipid Measurement (p = 0.06)	22 (71%)	31	78 (84%)	94
DM5: LDL Cholesterol Level (p = 0.23)	21 (96%)	22	69 (89%)	78
DM6: Urine Protein Testing (p < 0.001)	17 (55%)	31	80 (85%)	94
DM7: Eye Exam (p = 0.03)	12 (41%)	29	55 (62%)	89
DM8: Foot Exam (p = 0.13)	2 (7%)	28	15 (17%)	91

*N = Numerator; D = Denominator; % = Percent with measure

Although there are some EHR systems that capture diagnoses as text strings, the majority of commercial EHR systems capture diagnoses in coded form. In our study, diabetes was found on the problem list for 94% of the patients identified as having diabetes by expert review in the clinic studied. The problem list in EHRs has been shown to be about twice as accurate as problem lists maintained on paper,[17] although its reliability depends on the policies adopted by a clinic pertaining to problem list maintenance, and on the diligence with which these policies are followed. Building in benefits for EHR users based on reuse of diagnoses entered on the problem list (e.g., triggering reminders, providing decision-support) increase the motivation of users to keep the problem list up-to-date and accurate. EHR products should also allow for versioning of data within the EHR (e.g., problem lists, medication lists) so that the state of knowledge about an individual patient can always be reconstructed for a given point in time.

Calculating performance measures using only the subset of the target population determined by administrative data may significantly bias quality reports. In our study, patients with fewer than two face-to-face encounters for diabetes were significantly less likely to receive recommended diabetes care. For example, 97% of patients with two or more visits coded for diabetes had a glycohemoglobin measured within the preceding year, whereas only 68% of patients with known diabetes but seen fewer than two times for diabetes care during the study period had the test done on time. Since the encounter diagnosis is often triggered by tests ordered during the visit, using the encounter diagnosis as the inclusion criteria for the denominator effectively selects patients that are actively receiving care for diabetes, and amounts to a self-fulfilling prophecy. The study clinic produces its own internal clinical quality measures derived from the EHR developed specifically for quality improvement (which are not identical to the measures used in the DOQ project) and determined that it satisfied the clinical guideline 78% of the time across the entire population of seniors served. The clinic is managing its quality improvement programs using the internal clinical quality measure.

We recognize that limiting the target population to those with two coded visits may help exclude patients that may not be appropriate to include in the quality report for a specific clinic. For instance, patients with fewer than two diabetes visits may be obtaining diabetes care elsewhere. Unfortunately, this methodology may also exclude patients under the care of the practice who have untreated diabetes or unrecognized diabetes. Performance measures with an inadvertent bias that excludes patients with inadequately treated conditions may provide an inaccurate picture of the care actually provided. Some would argue, from a total quality perspective, that performance measures should also provide incentives for screening appropriate populations for a diagnosis, such as diabetes.

Another consideration when developing quality reports is whether to include patients in the denominator who may have just seen the physician at the tail end of the reporting period, offering no opportunity for appropriate intervention by the physician. We believe that a physician group is managing a population in addition to individual patients. Thus, outreach programs that increase awareness on the part of individuals with certain diagnoses or risk factors of the need for appropriate clinical attention may be an important responsibility of a health-care organization. Since health promotion and health-care delivery are community based services, one could argue that the performance measures for the services should also be community based. We acknowledge that there are differences of opinion about this assignment of responsibility.

Recognizing the importance of having robust measures of clinical performance in assessing quality and administering quality incentives, we are concerned about the systematic bias introduced by using metrics, such as those based on claims data, when more reliable clinical measures may be defined for practices using EHRs. Measuring performance can have positive effects on patient care, but accuracy is critical. Definitions that cause a systematic overestimate of quality delivered can cause organizational attention to be focused elsewhere, when, in fact, additional work needs to be done. Measures that underestimate the quality of care delivered can frustrate providers trying to improve, and deprive them of recognition they deserve. The most efficient solution is to reuse clinical data generated as a byproduct of clinical care. Ideally, data are entered once by the most appropriate professional for the purpose of providing care, and reused multiple times for the purposes of measuring quality, paying for performance, and generating knowledge about the effectiveness of treatments. Reuse of data not only improves data quality, it reduces the cost of secondary use of data—a welcome relief for providers often burdened with reporting mandates as an additional task or practice cost.

Use of administrative claims data as input for performance measures is also problematic for those provider organizations who are primarily capitated and do not need to collect billing data for their normal business operation. In this situation, it makes even more sense for quality data to be derived from EHR systems.

This study used manual review of the electronic health record as the gold standard. We did not identify nor contact the individuals involved to corroborate any of the findings contained in the documentation. Assessing the accuracy of the documentation was beyond the scope of this study.

Although it would be tempting to migrate national performance measurements to clinically based measures, currently only a minority of practices use EHR systems,[18] which would preclude its immediate use in performance measurement. It would be unfortunate, however, if, as the number of practices using EHRs grew, the country was tethered to a measurement system that lagged behind the deployment of clinical information systems. If pay-for-performance incentive programs continue to use performance measures derived from administrative data, it could have the unintended consequence of rewarding practices that did not convert to electronic systems (who report administrative quality measures with a systematic bias towards higher performance) or, alternatively, penalize those who report quality measures based on clinical data from EHR systems (which capture a larger denominator of patients). We believe policymakers should design measurement systems, and the incentive programs based on them, to take advantage of computer-based information systems that will become the new stan-

dard of care. Further study on this issue across a broader range of diagnoses should inform the development of clinically based quality measures. Likewise, a transition plan should be developed to migrate the nation's use of administratively based quality measures to clinically based quality measures. Given the inherent bias of claims-based quality measures to overstate performance, a temporary adjustment or premium should be built into incentives for reporting quality measures based on actual clinical data from an EHR system. This premium incentive could be time-limited to encourage more rapid adoption of EHR systems.

Developers of clinically based quality measures should take into account what data exist in coded form in EHRs. Efforts to standardize the relevant codes must be undertaken concurrently. Conversely, EHR products should be required to adopt data standards that support the measurement of standard quality measures.

Conclusions

While the small sample size limited the statistical power of our study, it does point out the urgent need to better understand the implications of using the two-encounter claims-based diagnosis rule before deciding on national standardized measures upon which to base public-reporting and reimbursement policies. Quality measurement organizations, the NQF, the Agency for Healthcare Research and Quality, and any coordinating body, such as the IOM's proposed National Quality Coordination Board,[19] should study this matter further to ensure that quality measures take advantage of physician-entered coded data in EHR systems. Most would agree that effective use of EHR systems holds great promise to systematically improve the quality of care provided to Americans.[20–23] Without adding burden to the care process, clinical data entered by clinicians into an EHR system at the point of care should be mined to generate new knowledge, measure performance, and reward those who deliver the best care with the best outcomes.

References ∎

1. Institute of Medicine. Crossing the Quality Chasm: A New Health System for the 21st Century. Washington, D.C.: National Academies Press; 2001.
2. Leape LL, Berwick DM. Five years after To Err Is Human: what have we learned? JAMA. 2005;293:2384–90.
3. Institute of Medicine Committee on Improving the Patient Record. The Computer-based Patient Record: An Essential Technology for Health Care, Revised Edition. 2 ed. Washington, D.C.: National Academies Press; 1997.
4. Tang PC, Lansky D. The missing link: bridging the patient-provider health information gap. Health Aff (Millwood). 2005; 24:1290–5.
5. Gans D, Kralewski J, Hammons T, Dowd B. Medical groups' adoption of electronic health records and information systems. Practices are encountering greater-than-expected barriers to adopting an EHR system, but the adoption rate continues to rise. Health Aff (Millwood). 2005;24:1323–33.
6. Institute of Medicine. Rewarding Provider Performance: Aligning Incentives in Medicare. Prepublication copy ed. Washington D.C.: National Academies Press; 2006.
7. Longo DR, Hewett JE, Ge B, Schubert S. The long road to patient safety: a status report on patient safety systems. JAMA. 2005 14;294:2858–65.
8. Jollis JG, Ancukiewicz M, DeLong ER, Pryor DB, Muhlbaier LH, Mark DB. Discordance of databases designed for claims payment versus clinical information systems. Implications for outcomes research [see comments]. Ann Intern Med. 1993;119:844–50.
9. Kennedy GT, Stern MP, Crawford MH. Miscoding of hospital discharges as acute myocardial infarction: implications for surveillance programs aimed at elucidating trends in coronary artery disease. Am J Cardiol. 1984;53:1000–2.
10. Iezzoni LI, Burnside S, Sickles L, Moskowitz MA, Sawitz E, Levine PA. Coding of acute myocardial infarction. Clinical and policy implications. Ann Intern Med. 1988;109:745–51.
11. Fisher ES, Whaley FS, Krushat WM, et al. The accuracy of Medicare's hospital claims data: progress has been made, but problems remain. Am J Public Health. 1992;82:243–8.
12. Peabody JW, Luck J, Jain S, Bertenthal D, Glassman P. Assessing the accuracy of administrative data in health information systems. Med Care. 2004;42:1066–72.
13. Benin AL, Vitkauskas G, Thornquist E, et al. Validity of using an electronic medical record for assessing quality of care in an outpatient setting. Med Care. 2005;43:691–8.
14. Fowles JB, Fowler EJ, Craft C. Validation of claims diagnoses and self-reported conditions compared with medical records for selected chronic diseases. J Ambul Care Manage. 1998;21:24–34.
15. Solberg LI, Engebretson KI, Sperl-Hillen JM, Hroscikoski MC, O'Connor PJ. Are claims data accurate enough to identify patients for performance measures or quality improvement? The case of diabetes, heart disease, and depression. Am J Med Qual. 2006;21:238–45.
16. Hebert PL, Geiss LS, Tierney EF, Engelgau MM, Yawn BP, McBean AM. Identifying persons with diabetes using Medicare claims data. Am J Med Qual. 1999;14:270–7.
17. Tang PC, LaRosa MP, Gorden SM. Use of computer-based records, completeness of documentation, and appropriateness of documented clinical decisions. J Am Med Inform Assoc. 1999;6:245–51.
18. Burt CW, Hing E. Use of computerized clinical support systems in medical settings: United States, 2001–03. Adv Data. 2005;353: 1–8.
19. Institute of Medicine Committee on Redesigning Health Insurance Performance Measures PaPIP. Performance Measurement: Accelerating Improvement. Washington, D.C.: National Academies Press; 2006.
20. Institute of Medicine. Patient Safety: Achieving a New Standard for Care. National Academies Press; 2004.
21. Institute of Medicine, Committee on Data Standards for Patient Safety. Key Capabilities of an Electronic Health Record System. 2003 Jul 31.
22. Bates DW, Kuperman G, Teich JM. Computerized physician order entry and quality of care. Qual Manag Health Care. 1994;2:18–27.
23. Bates DW, Gawande AA. Improving safety with information technology. N Engl J Med. 2003;348:2526–34.